Certifying the Uncertifiable: Aviation Safety Regulation for Learning Artificial Intelligence

Albert N. Clark

Independent Author

Published: August 28, 2026

ASX Research Journal and Database

ISSN 3068-3351 (Online)

Place of Publication: Cadiz City, Philippines

Publisher: ASXResearch.org

Author Note

Albert N. Clark

Department of Aerospace Sciences, ASXResearch.org

ORCID iD: https://orcid.org/0009-0002-7348-4395

The author reports no conflicts of interest.

Correspondence concerning this article should be addressed to Albert N. Clark,

Email: aclark@asxresearch.org

Abstract

Artificial intelligence is challenging one of aviation's most fundamental certification assumptions: that an approved system will behave tomorrow exactly as it did when it was certified. This paper examines the evolution of airborne automation into machine-learning-based systems and explores the regulatory, technical, operational, military, legal, and human-factors implications of certifying artificial intelligence whose behavior may be derived from data or modified through continued learning. Particular attention is given to emerging assurance concepts such as operational design domains, learning assurance, explainability, bounded adaptation, runtime monitoring, and lifecycle-based certification. Current developments involving Airbus, the Federal Aviation Administration, the European Union Aviation Safety Agency, the United States military, and DARPA demonstrate that increasingly autonomous and adaptive systems are already moving from theory into operational experimentation. The paper also considers the civil and legal ramifications of distributing decision-making authority among pilots, operators, manufacturers, software developers, and AI systems whose internal reasoning may not be immediately transparent. It concludes that the future of aviation certification may depend less on proving that every possible AI behavior is predetermined and more on demonstrating that adaptive behavior cannot escape a rigorously defined safety envelope. In this emerging framework, the central challenge is not whether aircraft can learn, but whether aviation can permit learning without surrendering accountability, predictability, and control.

Keywords: artificial intelligence, aviation certification, adaptive systems

Certifying the Uncertifiable: Aviation Safety Regulation for Learning Artificial Intelligence

Figure 1

Certifying the Uncertifiable: Aviation Safety Regulation for Learning Artificial Intelligence

Note. Click image for full-size image or click here.

Artificial intelligence did not arrive in aviation as a sudden replacement for the pilot; it emerged from a century-long progression in which machines were gradually entrusted with functions once performed exclusively by humans. The Sperry gyroscopic autopilot demonstrated automatic aircraft control in 1914, while subsequent generations of automatic flight control, flight management systems, autoland, terrain-warning systems, traffic collision avoidance systems, and fly-by-wire progressively transferred monitoring and control functions to software.

Yet these systems share an important characteristic: however sophisticated their algorithms may be, their intended behavior is engineered, bounded, tested, and essentially fixed before certification. Modern machine learning breaks with that tradition because its behavior is derived substantially from data rather than completely specified through conventional source-code logic. Demir et al. (2024), reviewing 224 aviation-safety studies, found artificial intelligence increasingly applied to accident analysis, pilot behavior, operational safety, prediction, and decision support. The historical progression therefore is not simply from manual flight to automation; aviation is approaching a transition from machines that execute instructions written by engineers to machines whose operational capabilities may be learned from data.

That distinction creates the central certification problem. Conventional airborne software certification rests on the proposition that engineers can define what a system is supposed to do, establish requirements, trace implementation to those requirements, verify the resulting software, and demonstrate that the approved configuration is the configuration installed in the aircraft. Machine-learning systems complicate every link in that chain. Their performance depends on training data, model architecture, statistical generalization, operational context, and conditions that may not have existed in the training set. Luettig et al. (2024) found that established aerospace system, software, and hardware standards are not fully applicable to machine-learning technologies and concluded that AI certification will require adaptations to traditional assurance practices. The difficulty becomes substantially greater if an operational model is permitted to learn after certification. In that case, the regulator is no longer merely being asked to certify complicated software; it is being asked to certify a process capable of changing the behavior of the software after the regulator has approved it. That is the paradox behind certifying the uncertifiable.

The aviation community is already constructing the intellectual machinery needed to address the problem. European research has been particularly influential through the concept of learning assurance, which extends conventional verification and validation into the data-driven development process. Pérez-Castán et al. (2022) applied this approach to a machine-learning air-traffic conflict detector and demonstrated an important weakness: performance metrics could change depending upon when predictions were made, meaning apparently acceptable aggregate accuracy could conceal operationally significant variation. EASA consequently developed a W-shaped assurance process that addresses data management, learning-process management, model training, model verification, implementation, independent verification, generalization, robustness, and explainability. Its framework also differentiates AI that assists humans from higher-authority human-AI teaming in which the machine can participate more directly in decision-making under human oversight. This distinction is critical because aviation certification cannot accept a claim that a neural network is 99.9% accurate as equivalent to demonstrating safety. A safety-critical aircraft system must also establish what happens during the remaining 0.1%, under what circumstances those failures occur, whether they are predictable, whether they can cascade into other systems, and whether another layer of protection prevents the error from becoming catastrophic.

The problem becomes even more difficult when the operating environment changes. A neural network trained to identify a runway, obstacle, aircraft, weather condition, maintenance anomaly, or conflicting trajectory can perform brilliantly against representative data while encountering something operationally novel tomorrow. Werner et al. (2026) argue that defining the operational domain and the AI constituent's Operational Design Domain is fundamental to learning assurance because certification requires boundaries around the circumstances in which

AI performance claims remain valid. The emerging philosophy is therefore remarkably similar to the way aviation already manages aircraft limitations: an AI system may ultimately be approved not because it can be proven infallible everywhere, but because engineers can establish the environment within which its performance is sufficiently assured. Outside that envelope, the system must recognize that it has reached its limit and transition safely to another system or a human operator. That could become one of the foundational principles of airborne AI certification: not artificial intelligence that knows everything, but artificial intelligence engineered to recognize when it no longer knows enough.

Aircraft manufacturers are not waiting for the philosophical debate to end. Airbus is explicitly developing artificial intelligence for autonomous-flight applications, including computer vision and machine-learning technologies intended to support navigation, obstacle detection, takeoff, and landing. More significantly for certification research, Airbus personnel are directly involved in scholarly investigation of machine-learning avionics qualification. Vidot et al. (2024) identify robustness, provability, and explainability as major obstacles to qualifying machine-learning-based avionics and describe potential applications ranging from vision-based navigation and obstacle sensing to autonomous flight, predictive maintenance, and cockpit assistance. Their work is especially significant because two of the authors are associated with Airbus while the research directly confronts the inadequacy of certification practices created for conventionally programmed software. The manufacturer question therefore has a definite answer: Airbus is tackling the problem, but the near-term objective is not an airliner that independently rewrites its flight-control intelligence between Paris and New York. The work instead concerns building the assurance architecture that could eventually make increasingly capable AI certifiable.

The military is moving considerably faster because its risk calculus, certification environment, and operational objectives differ fundamentally from commercial aviation. Delgado-Aguilera Jurado et al. (2024) conclude that military AI introduces risks that existing certification processes do not adequately address and argue for certification frameworks specifically adapted to trustworthy military AI. The theoretical problem is already colliding with real airplanes. DARPA's Air Combat Evolution program placed AI agents aboard the X-62A VISTA, a heavily modified F-16, and conducted autonomous within-visual-range combat maneuvering against a human-piloted F-16. The larger military objective extends beyond merely demonstrating that software can manipulate flight controls. It involves establishing trust, safety boundaries, human-machine collaboration, and autonomy capable of functioning in environments intentionally made unpredictable by an adversary. This distinction is enormously important for civil aviation research because military programs can expose adaptive autonomy to edge cases that would be unacceptable to explore initially aboard passenger aircraft. Lessons involving independent safety monitors, containment, human intervention, verification, and machine behavior under uncertainty could eventually migrate from military experimentation into civil certification philosophy.

The human remains the most complicated component in this architecture. Aviation history repeatedly demonstrates that automation can reduce workload and human error while simultaneously creating new categories of error through complacency, mode confusion, skill degradation, inappropriate trust, and delayed intervention. Grindley et al. (2026) found that appropriate trust in automation remains central to safe uncrewed-aircraft operation because both excessive intervention and excessive reliance can degrade performance. Yiu et al. (2026) similarly identify trustworthy and certifiable AI as central to the safe development of aviation

applications and emphasize human-AI collaboration as an important direction for continued research. The future cockpit problem is therefore not simply pilot versus AI. It is the allocation of authority between two fundamentally different cognitive systems. A machine can simultaneously monitor thousands of parameters without fatigue, yet encounter an edge case for which its training provides inadequate representation; a human may recognize the absurdity immediately but suffer fatigue, distraction, bias, or incomplete situational awareness. Certification will eventually have to demonstrate not only that the AI is safe and the human is competent, but that the relationship between them remains safe when they disagree.

Elon Musk provides an interesting boundary to the story precisely because his actual involvement should not be exaggerated. SpaceX has demonstrated extraordinarily advanced automated guidance, navigation, control, landing, rendezvous, docking, and recovery capabilities, and Musk's companies possess substantial artificial-intelligence expertise. However, there is presently no credible evidence that Musk, Tesla, xAI, or SpaceX is leading a program to certify adaptive machine-learning flight-control systems aboard civil transport aircraft. Claims that Musk is presently developing an AI-certified commercial airplane would therefore outrun the available evidence. His relevance is indirect but still useful: SpaceX demonstrates how far highly automated aerospace systems can be pushed when vehicle architecture, software, sensors, operations, and testing are developed as an integrated system. The distinction matters academically because automation is not synonymous with artificial intelligence, and artificial intelligence is not synonymous with online learning. A spectacular autonomous rocket landing can be accomplished by tightly engineered algorithms whose operational behavior remains constrained; it does not establish that an aircraft should be permitted to modify a safety-critical

neural network after certification. The temptation to conflate autonomy, AI, and adaptive learning is precisely the kind of conceptual shortcut that serious aviation research must resist.

The legal implications are enormous because aviation law is built around identifiable actors, products, duties, and standards of care. When a pilot makes an improper decision, investigators can examine training, procedures, judgment, fatigue, company policy, and regulatory compliance. When conventional equipment fails, investigators can trace design, manufacturing, maintenance, certification, and operation. A learning AI potentially distributes causation across the aircraft manufacturer, software developer, model developer, training-data provider, operator, maintenance organization, pilot, and even subsequent operational data that modified the system's behavior. Fadhil et al. (2025) examine this collision between explainable AI and legal accountability in aviation-related unmanned aircraft systems and emphasize the importance of traceability and explainability when responsibility must be established for AI-assisted safety-critical decisions. If an adaptive system learns a dangerous behavior after delivery, is the resulting accident a product defect, negligent operation, inadequate training, defective data governance, insufficient regulatory oversight, or some combination of them? The legal system will have considerable difficulty assigning fault to an algorithm unless aviation deliberately preserves the evidence necessary to reconstruct what the algorithm knew, what it inferred, why it acted, and how it had changed.

Civil aviation therefore faces a black-box problem far more profound than the conventional flight-data recorder. An accident investigator could theoretically recover an AI model and determine exactly what mathematical parameters existed at impact yet still struggle to explain why those parameters produced a particular decision. Explainable artificial intelligence is consequently not academic decoration; it may eventually become an element of airworthiness,

accident investigation, litigation, insurance, and due process. Degas et al. (2022) identified explainability as important to increasing AI acceptance in air traffic management, particularly as automated systems acquire greater influence over decision-making. The passenger environment adds another ethical dimension. Commercial passengers reasonably assume that an aircraft has been demonstrated safe before they board it; they have not consented to become training data for an evolving flight-control experiment. That reality strongly suggests that unrestricted real-time self-learning aboard passenger aircraft is unlikely to be the first regulatory destination. Controlled learning on the ground, carefully validated model updates, monitored operational feedback, rigorous version control, and regulator-approved deployment increments represent considerably more plausible intermediate steps.

This leads directly to what may eventually replace today's largely static certification model. Christensen et al. (2025) propose extending aviation's emerging W-shaped AI development framework with iterative engineering and operational feedback, allowing an AI or machine-learning constituent to be reevaluated as operational information accumulates. Their framework is particularly significant because it attempts to reconcile aviation's rigorous assurance culture with the iterative development philosophy common to contemporary software engineering. This points toward a potentially revolutionary regulatory concept: certification may cease to be solely an event and become, for certain AI functions, a controlled lifecycle. The regulator might approve the aircraft, initial model, training methodology, permissible data, learning boundaries, verification pipeline, rollback mechanism, and monitoring process. A new model could then be promoted into operational service only after automated and human assurance gates demonstrate that it remains within an approved safety envelope. Aviation would

not literally certify every future decision; it would certify a rigorously bounded mechanism through which future capability is allowed to evolve.

The most difficult frontier is true online learning: an aircraft that encounters new information in flight and modifies safety-critical behavior before landing. This is where the phrase certifying the uncertifiable becomes more than rhetoric. A regulator cannot meaningfully certify an unlimited universe of future neural-network states because the defining characteristic of such a system is that all future states are not known at the moment of approval. Certification could instead migrate toward establishing immutable boundaries around what the adaptive system is permitted to change. Those boundaries might encompass protected functions that cannot be modified, formally established safety constraints, independent runtime assurance, operational-domain restrictions, anomaly detection, immutable audit records, automatic reversion to a previously approved model, and human authority to disconnect or override the adaptive component. Luettig et al. (2024) emphasize precisely why conventional aerospace assurance practices encounter difficulty with machine-learning technologies: traditional processes were developed around systems whose behavior could be specified and verified in ways that data-driven models resist. The certification target may therefore eventually shift from proving that every future behavior is known to demonstrating that unknown future behavior cannot escape a known safety envelope.

The destination, then, is neither a sentient airliner nor the disappearance of pilots. It is potentially a new definition of airworthiness. Aviation spent more than a century building an extraordinary safety system around deterministic engineering, configuration control, redundancy, independent verification, human accountability, and the painful lessons preserved in accident investigations. Learning artificial intelligence challenges nearly every one of those pillars, but it

does not require abandoning them; it requires translating them. Demir et al. (2024) demonstrate how extensively artificial intelligence is already penetrating aviation-safety research, while Christensen et al. (2025) demonstrate how researchers are beginning to redesign engineering assurance itself around AI's fundamentally different characteristics. The safest future may involve aircraft that learn from enormous operational datasets, recognize hazards humans cannot perceive, predict failures before sensors cross conventional thresholds, assist controllers with traffic complexity beyond human cognitive capacity, and eventually perform increasingly autonomous flight. But the defining achievement will not be teaching an airplane to learn. Computer science has largely solved that part. The defining achievement will be teaching aviation how to permit learning without surrendering control of safety. If regulators, manufacturers, militaries, researchers, operators, lawyers, and pilots can solve that problem, aviation will have accomplished something more consequential than certifying another generation of avionics: it will have created a regulatory architecture capable of assuring systems whose capabilities can evolve after the engineers who designed them have stopped writing the instructions.

References

Christensen, J. M., Stefani, T., Anilkumar Girija, A., Hoemann, E., Vogt, A., Werbilo, V., Durak, U., Köster, F., Krüger, T., & Hallerbach, S. (2025). *Formulating an engineering framework for future AI certification in aviation.* Aerospace, 12(6), 482. https://doi.org/10.3390/aerospace12060482

Degas, A., Islam, M. R., Hurter, C., Barua, S., Rahman, H., Poudel, M., Ruscio, D., Ahmed, M. U., Begum, S., Rahman, M. A., Bonelli, S., Cartocci, G., Di Flumeri, G., Borghini, G., Babiloni, F., & Aricó, P. (2022). *A survey on artificial intelligence (AI) and eXplainable AI in air traffic management: Current trends and development with future research trajectory.* Applied Sciences, 12(3), 1295. https://doi.org/10.3390/app12031295

Delgado-Aguilera Jurado, R., Ye, X., Ortolá Plaza, V., Zamarreño Suárez, M., Pérez Moreno, F., & Arnaldo Valdés, R. M. (2024). *An introduction to the current state of standardization and certification on military AI applications*. Journal of Air Transport Management, 121, 102685. https://doi.org/10.1016/j.jairtraman.2024.102685

Demir, G., Moslem, S., & Duleba, S. (2024). *Artificial intelligence in aviation safety: Systematic review and biometric analysis*. International Journal of Computational Intelligence Systems, 17, 279. https://doi.org/10.1007/s44196-024-00671-w

Fadhil, T. H., Al-Haddad, L. A., & Al-Karkhi, M. I. (2025). *Legal accountability and UAV fault diagnosis explainable AI in aviation safety and regulatory compliance for liability challenges.* Discover Artificial Intelligence, 5, 410. https://doi.org/10.1007/s44163-025-00690-2

Grindley, B., Cherrett, T., Scanlan, J., & Plant, K. L. (2026). *Exploring the relationship between operator experience, propensity to trust automation and perceived system trustworthiness of uncrewed air vehicles.* Ergonomics. Advance online publication. https://doi.org/10.1080/00140139.2026.2625177

Luettig, B., Akhiat, Y., & Daw, Z. (2024). *ML meets aerospace: Challenges of certifying airborne AI*. Frontiers in Aerospace Engineering, 3, 1475139. https://doi.org/10.3389/fpace.2024.1475139

Pérez-Castán, J. A., Pérez Sanz, L., Fernández-Castellano, M., Radišić, T., Samardžić, K., & Tukarić, I. (2022). *Learning assurance analysis for further certification process of machine learning techniques: Case-study air traffic conflict detection predictor*. Sensors, 22(19), 7680. https://doi.org/10.3390/s22197680

Vidot, G., Gabreau, C., Ober, I., & Ober, I. (2024). *Qualification of avionic software based on machine learning: Challenges and key enabling domains*. Journal of Aerospace Information Systems, 21. https://doi.org/10.2514/1.I011164

Werner, F., Christensen, J. M., Stefani, T., Köster, F., Hoemann, E., & Hallerbach, S. (2026). *Formulating a learning assurance-based framework for AI-based systems in aviation.* Aerospace, 13(2), 200. https://doi.org/10.3390/aerospace13020200

Yiu, C. Y., Li, W. C., Ng, K. K. H., Chi, C. F., & Schiefele, J. (2026). *Enhancing aviation safety with artificial intelligence: A systematic literature review on recent advances, challenges and future perspectives*. Advanced Engineering Informatics, 71, 104378. https://doi.org/10.1016/j.aei.2026.104378